Glottal Source Parameter Estimation
by Comparison of Measured Signals with Simulated Signals

David G. Druker*, Ingo R. Titze*† and Brad H. Story†

*Department of Speech Pathology and Audiology, University of Iowa, Iowa City, Iowa 52242
†Wilbur James Gould Voice Research Center, Denver Center for the Performing Arts, Denver, Colorado 80204

Abstract: Electroglottograph (EGG) signals were digitized from a subject phonating a sustained /a/ vowel. A driven, kinematic glottal source model simulated the analogous physiologic signal—vocal fold contact area. Parameters controlling the source simulation were adjusted by a software optimizer so that the least-squared differences between the simulated and measured signals were minimized. These parameters included glottal adduction and convergence and effective vocal fold length, thickness and bulging during phonation. The optimized parameters are estimates, within the framework of this model, for the physiologic glottal configuration of the subject.

Speech simulation is the process of computing acoustic, and sometimes kinematic, signals from a model of the speech mechanism. Such simulation programs usually have two major components: a source model, which models the function of the vocal folds, and a vocal tract model. The source provides a pulsatile flow excitation to the tract, which computes wave reflection and propagation along its length and radiates an acoustic pressure from the mouth, and optionally, the nose. Taken together, source and tract models can have a large number of control parameters that determine the specific outputs. Accurately fitting measured data to these models requires adjusting several parameters simultaneously. Also, source models are highly nonlinear and can interact nonlinearly with the tract. So, in general, finding the parameters that best describe measured signals from a subject is a problem of nonlinear optimization (1).

The work presented in this paper is part of larger effort to build and test a techniques for estimating source and tract parameters from running speech. Such tools may be useful in speech-to-speech simulation and in understanding voice physiology. Here, an approach to solve the nonlinear optimization problem for only a driven kinematic source is described. The basic procedure is find the glottal model parameters minimizing the least-squares difference between measured EGG and simulated contact area.

SOURCE MODELS AND EGG

At least three major types of source models are in the literature. The simplest control glottal flow directly with several waveshape parameters. Fundamental frequency is specified directly. Typically these models are coupled to a vocal tract with no subglottal sections and receive no feedback from the tract. The source and tract together form a linear system of the form described in classic source-filter theory. One example is (2). Models of this type can produce realistic excitation to the tract, but do not inform us very much about physical behavior of the vocal folds. At the other extreme of complexity are self-oscillating models that emulate some of the tissue dynamics (3,4). These models are the most realistic available. However, they are difficult to control because the physical properties of live in situ tissue are not well known and the equations are sensitive to initial and boundary conditions.

The driven, kinematic model used in the present study is intermediate in complexity and fidelity to the real system. An earlier version is described in (5); the current version is summarized here. It calculates the spatial coordinates on the medial surfaces of the vocal folds as they vibrate and uses this information to find the minimum area of the glottal duct at each timestep. A rectangular coordinate system is erected with the origin on the midline at the posterior, inferior margin of the folds. Assuming the larynx is located in a body in standard anatomical position, the positive x-direction is medial, the positive y-direction is anterior, the positive z-direction is superior from the origin. The glottal width is

$$g(y, z, t) = 2(\xi_0(y, z) + \xi(y, z, t)).$$

where $$\xi_0(y, z)$$ is the prephonatory position along the x-axis and $$\xi(y, z, t)$$ is the vibrational displacement of the vocal folds. Glottal duct area is the minimum of

$$a(z, t) = \int_0^{h(z, t)} g(y, z, t) \, dy,$$

where $$h(z, t)$$ is the length of the glottal closure at time t. The model calculates the area of the glottal duct by integrating the width of the glottal gap along the length of the vocal fold. This process is repeated for each timestep, and the area is used to determine the effective vocal fold length, thickness, and bulging during phonation.

The parameters controlling the source simulation were adjusted by a software optimizer so that the least-squared differences between the simulated and measured signals were minimized. These parameters included glottal adduction and convergence, and effective vocal fold length, thickness, and bulging during phonation. The optimized parameters are estimates, within the framework of this model, for the physiologic glottal configuration of the subject.
where \( h(z, t) \) is the length of the glottis. Vocal fold contact area is

\[
a_c(t) = \begin{cases} 
\int_0^T h(z, t) \, dy, & \xi_0 \leq 0, \\
\int_0^T (L - h(z, t)) \, dy, & \xi_0 \geq 0.
\end{cases}
\]  (3)

where \( T \) is the thickness and \( L \) is the length of the folds while phonating. Since only contact area is used in this optimization, flow calculations will be ignored.

EGG measures the change in impedance of the neck due to vocal fold contact. It is not without artifact, but generally gives a good indication of the timing and characteristics of glottal closure (6).

**PROCEDURE AND RESULTS**

Glottographic and acoustic signals were digitized from a single subject performing a sustained /d/. Only the EGG signal was used in this study. A C++ program callable from the MATLAB environment, *pesto*, was used to simulate contact area. *Pesto* is based on work described in (4). Least-squares optimization, using the Levenberg-Marquardt method for search direction (1,7) was done in MATLAB.

At each iteration, *pesto* calculated the simulated contact area using parameters calculated from the previous step. One to seven parameters were optimized simultaneously. The simulated contact area was then normalized to size of the measured EGG. A cost (objective) function was calculated by subtracting a portion of one cycle of normalized contact area from a comparable section of measured EGG. After subtraction the part of the waveform that shows the closure event was weighted to twice the value of the rest of error signal. This was done because closure is the most significant acoustic event in the cycle.

The results showed good agreement between the simulated and measured signals. The residual (sum-of-squares) of the cost function for the best match was 0.478. Since the sum-of-squares (proportional to the average power) of the measured EGG signal was 33.57, it appears that simulated signal accounted for more than 98% of the variance of the measured signal. After optimizing, the sensitivity of the model around the solution was tested for the following parameters: glottal adduction and convergence and effective vocal fold length, thickness and bulging. In all cases, moving the parameter 10% of the physiological range away from the optimum increased the residual by at least 50%.

The goal of this experiment was to estimate glottal parameters. It is possible that mutual misadjustment of the parameters could produce a local minimum on the cost function. However, even with various starting points, the optimizer usually converged to the same parameter values. The glottal model never achieves perfect match with the measured signal. This is being investigated currently. Any inadequacies in the model will explain part of the difference. Also, most EGG devices either high-pass filter the output or apply an automatic gain control. Both of these would produce artifact that may be removable by signal processing.

**ACKNOWLEDGEMENTS**

This work was supported, in part, by Grant R01 DC02532 from the National Institute of Deafness and other Communication Disorders.

**REFERENCES**